Sequence-dependent sliding kinetics of p53

Leith, Jason S.; Tafvizi, Anahita; Huang, Fang; Uspal, William E.; Doyle, Patrick S.; Fersht, Alan R.; Mirny, Leonid A.; van Oijen, Antoine M.

Published in:
Proceedings of the National Academy of Sciences of the United States of America

DOI:
10.1073/pnas.1120452109

Publication date:
2012

Sequence-dependent sliding kinetics of p53

Jason S. Leith[a], Anahita Tafvizi[b], Fang Huang[c], William E. Uspal[d], Patrick S. Doyle[d], Alan R. Fersht[c], Leonid A. Mirny[e,1,2], and Antoine M. van Oijen[f,1,2]

[a]Program in Biophysics, Harvard University, Cambridge, MA 02138; [b]Department of Physics, Harvard University, Cambridge, MA 02138; [c]Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom; [d]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Institute for Medical Engineering and Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139; and [f]Zernike Institute for Advanced Materials, University of Groningen, 9747 AG Groningen, The Netherlands

Edited by Gordon G. Hammes, Duke University, Naples, FL, and approved August 16, 2012 (received for review December 15, 2011)

Proper timing of gene expression requires that transcription factors (TFs) efficiently locate and bind their target sites within a genome. Theoretical studies have long proposed that one-dimensional sliding along DNA while simultaneously reading its sequence can accelerate TF's location of target sites. Sliding by prokaryotic and eukaryotic TFs were subsequently observed. More recent theoretical investigations have argued that simultaneous reading and sliding is not possible for TFs without their possessing at least two DNA-binding modes. The tumor suppressor p53 has been shown to slide on DNA, and recent experiments have offered structural and single molecule support for a two-mode model for the protein. If the model is applicable to p53, then the requirement that TFs be able to read while sliding implies that noncognate sites will affect p53's mobility on DNA, which will thus be generally sequence-dependent. Here, we confirm this prediction with single-molecule microscopy measurements of p53's local diffusivity on noncognate DNA. We show how a two-mode model accurately predicts the variation in local diffusivity, while a single-mode model does not. We further determine that the best model of sequence-specific binding energy includes terms for "hemi-specific" binding, with one dimer of tetrameric p53 binding specifically to a half-site and the other binding nonspecifically to noncognate DNA. Our work provides evidence that the recognition by p53 of its targets and the timing thereof can depend on its noncognate binding properties and its ability to change between multiple modes of binding, in addition to the much better-studied effects of cognate-site binding.

protein-DNA interactions | protein-DNA search | promoter search | energy landscape | one-dimensional diffusion

Tumor suppressor p53 is known as the "guardian of the genome." The protein is mutated in more than 50% of cancers (1), and plays important roles in activating DNA-repair, cell-cycle arrest, and apoptosis. To prevent the replication of damaged DNA, damage-activated p53 must reach its target promoters sufficiently fast.

In addition to its clinical importance, p53 is the first eukaryotic transcription factor (TF) directly observed to undergo one-dimensional (1D) diffusion on DNA (2). This 1D sliding has long been proposed to facilitate search by DNA-binding proteins (DBPs) (3, 4) and was characterized by single-molecule experiments for RNA polymerase (5, 6), mitochondrial repair enzymes (7), *lac* repressor (8, 9), and repair complex Msh2–Msh6 (10) (see ref. 11 for review). Sliding of p53 was first demonstrated by bulk biochemical experiments and was shown to play a role in p53's activation of target genes (12, 13). Several theoretical and experimental studies (7, 8, 14, 15) have shown that despite a vast excess of accessible DNA (10^7–10^9 bp) to which DBPs have nonspecific affinity, their search process can be efficient if they alternate rounds of 1D sliding while bound nonspecifically with rounds of three-dimensional (3D) diffusion between different sections of DNA. Until recently, in vivo studies have been limited to bacterial systems (8), and it remained unclear whether the same mechanism was at play in eukaryotes where DNA is packed by nucleosomes limiting space for sliding. Recently Larson et al. (16) have demonstrated that yeast DBPs search for their sites by a 1D/3D mechanism. By demonstrating the ability of p53 not only to slide but also to "read" DNA while sliding, our study provides strong support for a 1D/3D mechanism in high eukaryotes. An earlier single-molecule study of Harada et al. (6) found sequence-dependence in the dissociation kinetics of RNA polymerase, but not in its sliding kinetics.

For sliding to be functional in facilitating search, DBPs must be able to read the DNA sequence they slide on. This implies that the binding energy at each DNA position depends on the sequence. The magnitude of this sequence dependence can be captured by the standard deviation of the energies comprising the landscape, σ. Prior theoretical work (17) demonstrated that $\sigma \lesssim 1.5\ k_B T$ is required for fast sliding and facilitated search, while stability of the protein–DNA complex requires $\sigma \gtrsim 5\ k_B T$. These mutually unsatisfiable requirements lead to an apparent "speed-stability paradox"(17, 18), which had been qualitatively anticipated (19).

A multimode model protein–DNA interaction has been proposed to resolve this paradox (17, 18): In their simplest form, DBPs exhibit two modes of binding and stochastically switch between them: a search (**S**) mode, and a recognition (**R**) mode (Fig. 1 *A* and *B*). In **S** mode, DBPs bind with small sequence-dependence ($\lesssim 1.5\ k_B T$) and slide efficiently. In **R** mode, DBPs bind highly specifically and sliding is negligible. Fast search requires the **S** mode to have significantly lower average energy and thus be favored at nearly all DNA positions to avoid unproductive visits to **R** mode (18). Transition into **R** mode at cognate-like sites that have low energy slows down sliding, but allows recognition of the cognate site. The central idea of this study is that such slowdown on near-cognate sites can be detected experimentally.

Experimental evidence supporting this multimode model includes NMR studies that found specific and nonspecific conformations for *lac* repressor (20, 21) on DNA, and X-ray crystallography of the restriction enzyme BstYI (22). While an NMR study of homeodomain hopping and sliding on DNA found only small structural changes between specific and nonspecific binding, the homeodomain is only weakly specific, with only 200-fold preference for its cognate site over nonspecific DNA (23), compared with p53's $\sim 10^5$-fold preference (24). Transcriptional activation by the yeast TF *Mbp1* has been shown to involve 1D sliding (16). Recently, multiple conformations have been identified by electron microscopy for p53-DNA complexes (25). Further support for a multimode search comes from a single-molecule study of p53 truncation mutants showing that distinct

Author contributions: J.S.L., L.A.M., and A.M.v.O. designed research; J.S.L. performed research; J.S.L., A.T., F.H., W.E.U., P.S.D., and A.R.F. contributed new reagents/analytic tools; J.S.L. analyzed data; and J.S.L., L.A.M., and A.M.v.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]L.A.M. and A.M.v.O. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: leonid@mit.edu or a.m.van.oijen@rug.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120452109/-/DCSupplemental.

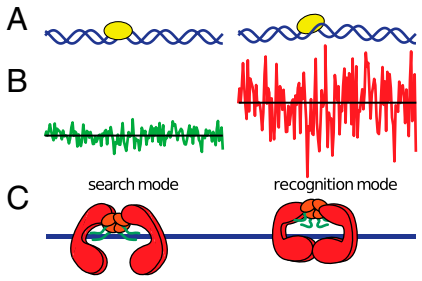


Fig. 1. Energy landscapes and cartoons of proteins on DNA in search (**S**) and recognition (**R**) modes. (*A* and *B*) In **S** mode, a generic protein (yellow) interacts chiefly with the DNA backbone and experiences a smooth landscape. In **R** mode, it interacts with the nucleobases, yielding a highly sequence-dependent landscape. (*C*) Cartoon model for p53, based on EM data (25), indicates the domains responsible for the modalities: green C-terminal domain for the **S** mode; red core domain for the **R** mode. Tetramerization domain in orange.

p53 domains (C-terminal and core) perform sliding and recognition functions (26) (Fig. 1*C*). The C-terminal domain's energy landscape is estimated to have $\sigma \approx 0.6\ k_B T$, satisfying the requirements for efficient search, while the specifically-binding core domain cannot slide on its own.

Here, we report measurements using single-molecule fluorescence microscopy of p53's sequence-dependent diffusivity. We observe that p53's sliding kinetics on λ-phage DNA in the absence of known cognate sites vary by a factor of 1.6 among different regions of DNA. Using a model with both **R** and **S** modes and a model with only a single mode, we construct predicted effective energy landscapes for p53 on DNA and demonstrate that the two-mode model but not the one-mode model accounts for the observed variation in diffusivity among regions of the DNA. We further provide evidence that the two identical homodimers making up biologically active tetrameric p53 can bind DNA in dissimilar modes (i.e., "hemi-specifically"). Such binding has been observed for other DBPs binding to oligonucleotides containing "half-sites"(22). Our analysis of p53's sequence-dependent sliding kinetics reveals that hemi-specific binding is widespread in p53-DNA interactions.

Results

To assess whether p53's diffusivity depends on its position on λ-phage DNA, we recorded trajectories of fluorescently labeled single p53 molecules on DNA that was tethered to the surface of a flow cell and stretched by shear flow, using total internal reflection fluorescence microscopy (Fig. 2). We mapped trajectories to the contour of DNA (Fig. 3*A*) and made maximum likelihood estimates of diffusion coefficients, D, of p53 particles, while accounting for drift from buffer flow and position-dependent DNA fluctuations (*Data analysis* in *Methods*). We found that DNA fluctuations cannot account for observed particle diffusivity: The square of the central 95% of the range of the p53 particles exceeds the amplitude of the square displacement of probes covalently attached to the DNA in the limit of long time windows, Δt, by 1–3 orders of magnitude (Fig. 3*B*). Each p53 trajectory's D, along with the particle's range covered on the DNA, is shown in Fig. 3*C*.

We observed that different regions of the λ-phage DNA correspond to different diffusion coefficients. We determined an aggregate experimental diffusion coefficient, D_{expt}, for each segment by assigning every midpoint of each particle trajectory displacement to a position on the DNA, binning the contour of the DNA into approximately 3-kb segments, and calculating the mean D within each segment (Fig. 3 *A*, *C*, and *D*). Error bars in Fig. 3*D* are errors of estimated D_{expt} calculated as standard deviation in D from 1,000 bootstrap resamples of the particles in each segment (*SI Text*, *Significance and Consistency of Experimental Results*). We found that segments' aggregate D spanned a factor of approximately 1.6, with 11 of the 36 pairs of segments differing in D significantly at $\alpha = 0.05$ and 6 of the pairs at $\alpha = 0.01$ (*SI Text*, *Significance and Consistency of Experimental Results*). Plots of the mean-square displacement as a function of Δt for particles in selected segments are shown in Fig. 3*E*.

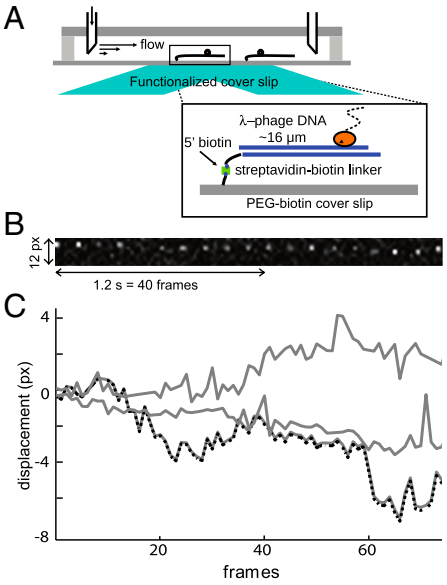


Fig. 2. Experimental setup and initial data analysis. (*A*) Biotinylated λ DNA is flowed into the cell and adheres to the streptavidin-coated surface. The DNA is stretched by hydrodynamic drag. Labeled p53 proteins are imaged diffusing along the DNA. After a series of protein movies are taken, the DNA is stained and imaged. (*B*) Kymogram of a single p53 protein diffusing on DNA. Flow direction is up; every fourth frame is shown, giving an apparent frame rate of 120 ms. (*C*) Trajectories of three particles (gray). The dotted black trace represents the bottommost trajectory (kymogram in *B*), corrected for drift.

Next, we tested whether this variation in D_{expt} could be explained by sequence-specific binding in **R** mode. Because λ-phage DNA contains many sites that resemble half- and full-sites of p53, we expected p53 could bind these sites and thus slow down sliding. To this end, we developed a model of an effective two-mode landscape experienced by tetrameric p53 on DNA. Experimental studies have demonstrated that p53, a dimer of dimers with a response element (RE) of 20 bp, binds with one of its dimers to a 10-bp half-site with greater affinity than to random DNA (24, 27). Accordingly, we posited that each dimer could bind a position on DNA in **R** mode with an energy depending on the sequence, $E_R(x)$, or in **S** mode with constant energy E_S, while the other dimer could bind in a similar (fully specific or fully nonspecific) or dissimilar mode (hemi-specific) (*Prediction of diffusion coefficients* in *Methods*). A cooperativity term ε accounts for additional binding energy when both dimers bind in **R** mode.

We built a sequence-specific landscape $E_R(x)$ using a position weight matrix (PWM) for a single dimer, based on known p53 REs (28) (Fig. 4*A*). The PWM was robust to the literature source of REs (*SI Text*, *Prediction of Energy Landscape and Local Diffusion Coefficients*). Then, an effective binding energy for the tetramer, $U(x)$, was calculated over all positions according to the two-mode model (Eq. **4** in *Methods* and *SI Text*, Eqs. **S15–S18**), and also according to a single-mode model, giving rise to respective energy landscapes (Fig. 4 *C* and *E*). The calculation of $U(x)$ allows variable spacing between the 10-bp sequences bound by the two dimers. We identified sites of λ DNA that scored as well as some the weaker known p53 REs, but are not known to be in vivo targets of p53 (Fig. 4*B*). The PWM was scaled to fit experimentally measured dissociation constants of p53 and oligonucleotides

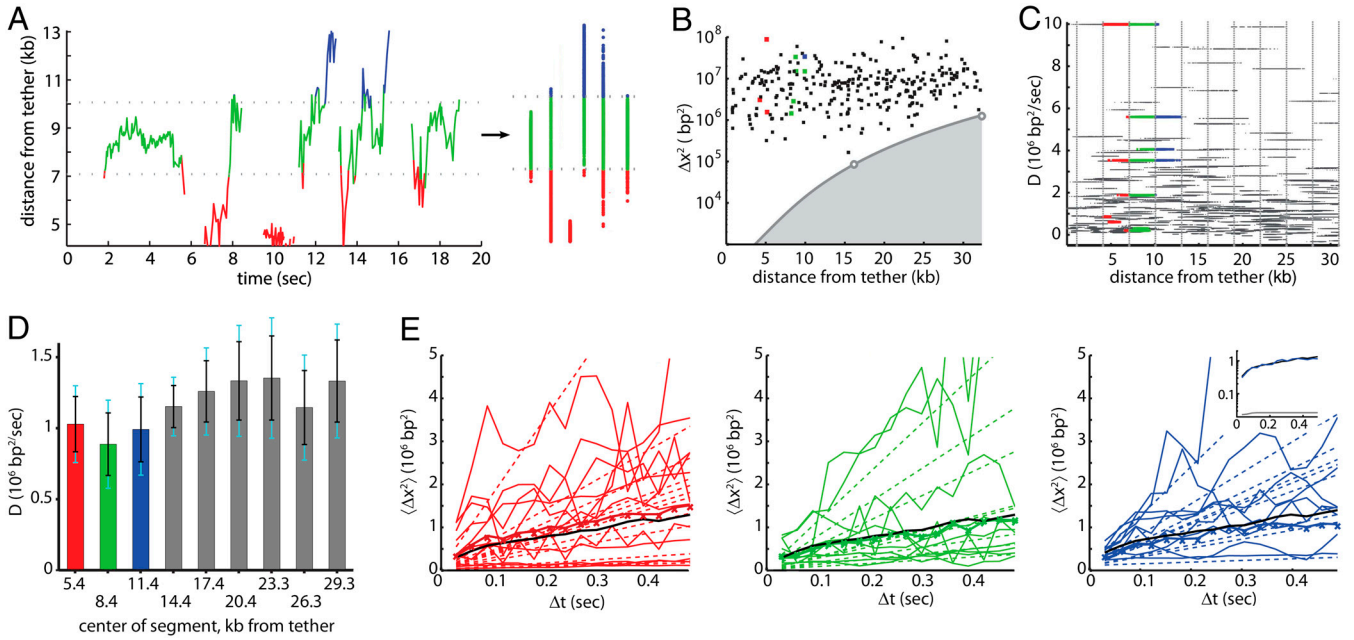


Fig. 3. Data analysis: diffusion coefficients of p53 on λ-phage DNA. (*A*) Trajectories of selected particles in three representative segments. The trajectories have been spread out horizontally for clarity. Portions of trajectories are colored according to the segment in which they lie: red, green, and blue for segments 1, 2, and 3, respectively. The positions and assigned segments for each particle's displacements are shown to the right. (*B*) Squares: squared range of the central 95% of each trajectory, plotted over the trajectory's midpoint. Colored squares represent particles shown in *A*. Gray circles: mean squared displacement (MSD) at long (>100 ms) time windows, Δt, of quantum dots fixed at one-third and two-thirds the contour length from the tether point. Shaded region is the MSD of the DNA at long Δt. (*C*) Horizontal lines consist of dots plotted on the horizontal axis at the midpoint of each displacement within a trajectory, and on the vertical axis at the *D* corrected for drift and DNA fluctuations of their respective particle. Colored dots correspond to colored dots in *A*. (*D*) Estimated D_{expt} for each segment, from averaging all values in *C*. Colored bars correspond to coloring scheme for *A*–*C*. Uncertainties were determined by bootstrapping: The particles contributing to each segment's *D* were resampled 1,000 times, and the resulting diffusion coefficients calculated. Black error bars represent a standard deviation in the resampled diffusion coefficients above and below the estimated *D*; cyan error bars sample only half the number of particles (*SI Text*, *Significance and Consistency of Experimental Results*). (*E*) Thin solid traces are MSD/Δt for particles whose median position lies within segments 1 (red), 2 (green), or 3 (blue). For clarity, only every third particle is shown. Each particle is analyzed as in *Methods*, and for each particle shown, a corresponding dashed line plotted with slope 2*D*. Thick solid colored traces are the weighted MSD/Δt for the particles shown. Solid black traces are the MSD/Δt for all particles in all nine segments presented. Inset (same axis units) shows MSD/Δt for the DNA (gray); black and thick blue lines are the same as in the main panel.

containing full-sites, half-sites, and random sequences (24). The difference between E_S and the mean of $E_R(x)$ was set so that the free-energy difference between specific and non-specific binding for our landscape would match that for typical eukaryotic TFs (29).

From the computed landscape, we predicted each segment's reduction in diffusivity relative to a flat **S** landscape, D/D_0 (*Methods* and *SI Text*, *Prediction of Energy Landscape and Local Diffusion Coefficients*). Areas with more/deep energy wells were found, as expected, to correspond to a reduced diffusivity of particles in

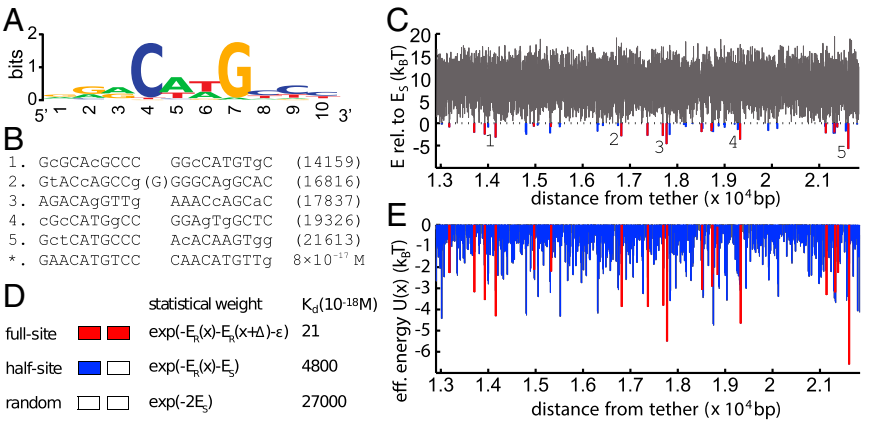


Fig. 4. Theory: scoring the λ genome and predicted landscapes. (*A*) Half-site sequence logo for p53. (*B*) 1–5: Sequences and positions in bp from the tether of full-sites found in segments 4–6 of λ DNA shown in *C* and *E*. Lowercase letters indicate nucleotides that do not match the consensus sequence of RRRCWWGYYY. Sequence 5 is the strongest-scored fullsite among all segments. Asterisk (*): Sequence from among the known p53 RE whose K_d has been measured by Weinberg et al. (40) that has the shortest Hamming distance to sequence 5. (*C*) Predicted one-mode landscape in segments 4–6. Red, full-sites; blue, half-sites. (*D*) Key elements of two-mode model. The statistical weights of fully specific, hemi-specific, and nonspecific binding at a position *x* in making up $U(x)$ [**5**] are indicated. For the great majority of positions on DNA that lack a half-site, the greatest of these is the term representing fully nonspecific binding. For positions that include a half-site, it is the term representing hemi-specific binding. For full-sites, it is the term representing fully specific binding. K_d's are of p53 to representative of full-sites, half-sites, and nonsites (24). (*E*) Predicted two-mode effective landscape. Most positions are dominated by nonspecific binding. The possibility of hemi-specific binding makes half-site binding relatively more important for the two-mode model than for the one-mode model.

these areas (Fig. 5*A*). We demonstrate that D/D_0 is ratio of the time t_s a protein spends sliding in **S** mode to the total time the protein spend on the landscape, t_{total}:

$$\frac{D}{D_0} = \left\langle \frac{t_s}{t_{\text{total}}} \right\rangle = \frac{n \exp(-2E_S)}{\sum_x^n \exp(-U(x))}. \quad [1]$$

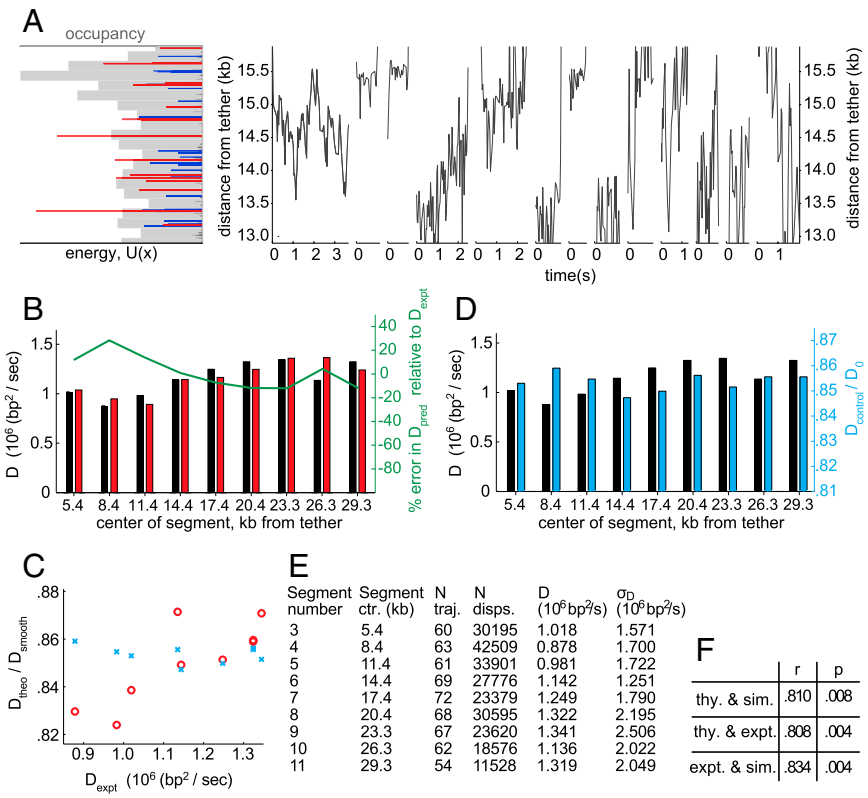
This result is based on the assumption that traps are isolated or that the protein does not slide in **R** mode.

We compared D_{expt} with D/D_0 over the segments and found the experimental and predicted diffusion coefficients to correlate strongly ($r = 0.81$, $p = 0.008$) (Fig. 5*B*, black and red bars). To assess the significance of this correlation, r_{expt}, we constructed 500 random permutations of $E_R(x)$, computed $U(x)$ and D/D_0 for each of them, and determined the resulting r_{ctl} between predicted D/D_0 and experimental D_{expt} over the segments (example in Fig. 5 *C* and *D*). The correlation between predicted D/D_0 and D_{expt} exceeded that between control D/D_0 and D_{expt} for all but 4 of the 500 control landscapes ($p = 0.008$). The observed strong and significant correlation demonstrates that a two-mode sequence-specific landscape can explain the observed positional variability of p53's diffusion coefficient. Summary statistics of the segments' D_{expt} are given in Fig. 5*E*.

We also compared the ruggedness of effective landscape $U(x)$, formed from an **R** and an **S** landscape, with earlier experiments and with theoretical requirements. Satisfyingly, the global ruggedness σ of the two-mode landscape is 0.51 k_BT, which lies below the theoretical upper limit for efficient search, ~1.5 k_BT, and falls within the uncertainty for the aggregate σ, 0.84 ± 0.40 k_BT obtained for p53 earlier (2). In contrast, the landscape without a nonspecific binding mode has σ = 3.5 k_BT, which is too great on theoretical grounds for efficient sliding and moreover is incompatible with observed diffusion coefficients. Furthermore, the diffusivity D/D_0 computed for the one-mode landscape shows no significant correlation with D_{expt} ($r = 0.51$, $p = 0.11$), ruling out the one-mode model.

Experimental observations of dimeric p53 binding to half-sites (27) prompted us to explore hemi-specific binding by tetrameric p53. The two-mode model discussed thus far does not require that the two dimers making up the tetramer bind in the same mode. This enhances p53's affinity even to those half-sites that are flanked by sequences that would be unfavorable to bind in **R** mode (Fig. 4*D*). When we eliminated from our model the hemi-specific mode—represented by the two middle terms of Eq. **4**, the correlation between segments' D_{expt} and predicted D/D_0 decreased ($r_{\text{no hemi}} = 0.72$ versus $r = 0.81$). The fraction of the sequence-specificity of p53's diffusion coefficient that owes to full-sites is thus approximately $(r^2_{\text{no hemi}})/r^2 = 0.78$, and the fraction that owes to half-sites is approximately 0.22.

As a further test of the two-mode model, we performed Gillespie simulations (30) of a protein undergoing a random walk on the two-mode landscape. Drift from buffer flow and noise



E

Segment number	Segment ctr. (kb)	N traj.	N disps.	D (10^6bp^2/s)	σ_D (10^6bp^2/s)
3	5.4	60	30195	1.018	1.571
4	8.4	63	42509	0.878	1.700
5	11.4	61	33901	0.981	1.722
6	14.4	69	27776	1.142	1.251
7	17.4	72	23379	1.249	1.790
8	20.4	68	30595	1.322	2.195
9	23.3	67	23620	1.341	2.506
10	26.3	62	18576	1.136	2.022
11	29.3	54	11528	1.319	2.049

F

	r	p
thy. & sim.	.810	.008
thy. & expt.	.808	.004
expt. & sim.	.834	.004

Fig. 5. Comparison of theory, simulations, and experiment. (*A*) (*Right*) Random third of trajectories with center in segment 4, ordered by increasing estimated *D*. (*Left*) Predicted potential wells denoted by red (full-sites) and blue (half-sites) bars, with height of bars proportional to predicted effective energy, $U(x)$. Gray bars are a histogram of observed occupancy of all particles within the segment, with bin widths equal to one-twentieth the segment width. (*B*) Estimated *D* for experimental (black bars, same as in Fig. 3*D*) and predicted D/D_0 (red bars) for the predicted landscape, over segments along λ DNA. D/D_0 is scaled to match D_{expt}'s mean and coefficient of variation. Green trace is the percent error in predicted D/D_0 normalized by the mean D/D_0, relative to D_{expt} normalized by the mean D_{expt}. (*C*) Scatter plot of D_{expt} versus D/D_0 for all segments. Red circles correspond to values for the predicted landscape based on the two-mode model; cyan x's correspond to values for the control landscape whose correlation with D_{expt} was the median from among the 500 control landscapes. (*D*) Black bars are identical to those in *B*. Cyan bars correspond to D/D_0 of the control landscape that produced the cyan x's in *C*. (*E*) Summary statistics of experimental data. For each segment, we show the number of particles contributing to the segment's D_{expt}, the number of displacements contributing likewise, the estimates of D_{expt} as determined in *Data analysis* in *Methods*, and the standard deviation of particle's *D* weighted by the number of displacements contributed by that particle. (*F*) Correlation coefficients and *p*-values.

from DNA fluctuations were added to the simulated data, which we analyzed identically to the data for experimental Ds. The simulated and experimental diffusion coefficients across the segments correlate strongly ($r = 0.834$, $p = 0.004$) (Fig. 5*F*). Simulations on the same 500 control landscapes described above were used to determine statistical significance, and thus provide similar validation as do analytical results of the sequence-specific sliding of p53 by the mechanism of two modes of interaction with DNA.

Discussion

We previously proposed a two-mode model of protein–DNA interaction that allows for fast search and specific binding (17, 18). Our earlier single-molecule measurements of p53 sliding on DNA revealed that the protein slides with sufficiently low friction to satisfy the model's requirements for efficient search (2). The present study shows that p53 can read the sequence of the DNA on which it is sliding, which is essential for sliding to be functional in accelerating target localization. Our data further suggest that the protein reads, in addition to canonical and near-canonical 20-bp full-sites, half-sites of 10 bp. This conclusion agrees with recent studies showing a role for hemi-specific binding in transcriptional activation at high p53 expression levels (31).

Our results indicate that hemi-specific binding is a general phenomenon of p53-DNA interactions, and not limited to a few known half-site REs. In addition to transcriptional activation, we conjecture that hemi-specific binding might serve to titrate p53 or bias the preactivation distribution of p53 on DNA. This latter function especially is suggested by clustering of degenerate p53 REs near canonical REs, which has also been found for other mammalian TFs (32). Odd numbers of half-sites have been found in many p53 binding sites (33, 34); hemi-specific binding would allow finer tuning of transcriptional activation of p53's targets.

Our model of p53-DNA binding energy is based on a PWM approximation that was shown to be sound for the four eukaryotic TFs studied (29) as well as for p53 in the case of good full-sites (35), but it omits some observed peculiarities of p53 REs such as gaps within half-sites(33), stronger conservation within a full-site of the first half-site than the second (33), and transcriptional activation from three-quarter sites (36). Accounting for these complexities might yield a stronger correlation between predicted and experimental D, at the expense of model simplicity.

The two-mode model can be generalized to include transition states or a reaction coordinate of the conformational transition in the protein–DNA complex (17, 18). Molecular dynamics studies of TF–DNA association indeed show a range of conformations (37). Our estimate of D/D_0 is equivalent to the ratio of partition functions of a flat landscape and the predicted "golf-course" landscape. As such, it is independent of the transition rates between **R** and **S** modes, since holding the binding energy of the protein in **R** mode and in **S** mode at a given position constant requires the **R**-to-**S** and **S**-to-**R** rates to vary by the same constant factor. On sufficiently long timescales of sliding, a visit to the **R** state that lasts n times longer will happen n times less frequently. And although our multiple landscapes are motivated by our prior work on the global protein–DNA search, they are compatible with other such treatments as well (38, 39).

While enough sites on the λ genome have sufficient affinity for p53 to measurably affect its sliding kinetics, in future work, we intend to add known p53 REs to λ and other DNA (*SI Text*, *Control for Specific Binding*) to observe stronger sequence-dependent effects, and to answer questions such as whether p53 can slide over a target site without recognizing (binding stably to) it.

We report here the observation of sequence-dependent 1D diffusional kinetics of p53 on DNA. We offer additional experimental support for the importance of 1D diffusion in the kinetics of transcriptional regulation and protein–DNA recognition. With p53 at least, a full understanding of how its complex promoter architecture functions in transcriptional regulation requires consideration of moves by the protein on DNA even after it has found its cognate site and the ability of the protein to recognize both full- and half-sites while undergoing those moves. Evidence for a multimode model of p53's binding to DNA suggests that the protein's function may be disrupted not only by the comparatively well-studied mutations in cognate-site binding residues, but also by mutations that affect its nonspecific interaction with DNA or its ability to transition between specific and nonspecific modes, with potential importance for human health.

Methods

Materials and Data Acquisition. The optical setup, DNA constructs, labeled p53, and flow cells (Fig. 2*A*) were as described previously (2), excepting that the protein was labeled with AlexaFluor 555 (Invitrogen) and illuminated by the 532 nm line of a Nd:YAG laser, and that fiduciary beads were used to align p53 movies and DNA movies (*SI Text*, *Materials and Data Acquisition*).

Data Analysis. Protein molecules were assigned to individual DNA molecules, and their trajectories were recorded using scripts written in MATLAB (MathWorks). Positions of the p53 molecules in space along the DNA image were mapped to positions on the contour of the DNA using Brownian dynamics simulations (*SI Text*, *Data Analysis*). Fig. 2*C* shows three sample trajectories.

We determined a diffusion coefficient D for every trajectory recorded except for those of particles stuck to the flow-cell surface (*SI Text*, *Materials and Data Acquisition*). We used maximum likelihood estimation, correcting for biased drift from buffer flow as well as for fluctuation in the λ-phage DNA on which p53 diffused. We found D for a particle to be estimated by

$$D = \frac{1}{2}\left(\frac{1}{n}\sum_{i}^{n}\frac{(\Delta x_i - v\Delta t_i)^2}{\Delta t_i} - \frac{1}{n}\sum_{\Delta t} n_{\Delta t}\frac{\langle \Delta x^2_{d,\Delta t}\rangle}{\Delta t}\right), \qquad [2]$$

where v is the drift velocity (*SI Text*, Eq. **S3** in *Data Analysis*).

An N-frame trajectory contains $(N-1)(N-2)/2 \equiv n$ displacements. The ith observed displacements in space and in time are, respectively, Δx_i and Δt_i. The second sum in Eq. **2** is over time windows Δt ranging from the camera frame rate, 30 ms, to 2 s. The quantity under the sum is the mean squared displacement of the DNA itself owing to Brownian fluctuations, $\langle x^2_{d,\Delta t}\rangle$ on a timescale of Δt, as calculated from measurements of quantum dots covalently attached to the DNA (*SI Text*, *Interpolations of DNA-Fluctuation Variance and Distributions*), divided by Δt, and weighted by the number of displacements with a corresponding Δt. The first sum represents the apparent diffusion coefficient of p53, corrected for drift. D_{expt} for a segment is the average of all particles in the segment, weighted by the number of displacements contributed.

Prediction of Diffusion Coefficients. In its most basic form, our two-mode model posits a recognition (**R**) mode in which a protein's binding energy, E_R, is dependent on its position on DNA, x, and a search (**S**) mode in which its binding energy, E_S, is constant.

To determine $E_R(x)$, we scored the λ genome using position weight matrices (PWM_S) of p53 half-sites derived from a catalogue of p53 binding sites (28) (Fig. 4*A*). We assume that differences between scores are proportional to differences between corresponding half-site energies:

$$E_R(x) - E_S = c(PWM(x) - PWM_S), \qquad [3]$$

where $PWM(x)$ is the score for position x, and PWM_S is the score corresponding to binding energy in the **S** mode, and c is a conversion factor between score and energy (in k_BT). Since p53 is a dimer of dimers with each dimer able to bind independently to a 10-bp half-site (27). For binding in **R** mode, then, the left dimer binds with energy $E_R(x)$ and the right dimer binds with energy $E_R(x+\Delta)$. The statistical weight $w(x)$ of a site x is thus the sum of the Boltzmann factors corresponding to each of the four modes, and the negative logarithm of $w(x)$ is the effective binding energy of p53 to λ DNA at position x:

$$U(x) = -\log w(x);$$
$$w(x) = e^{-2E_S} + e^{-(E_S+E_R(x+\Delta))} + e^{-(E_R(x)+E_S)} + e^{-(E_R(x)+E_R(x+\Delta)+\varepsilon)}, \quad [4]$$

with the spacing between the 5' ends of half-sites, Δ, usually a half-site length, 10 bp. It can be greater for full-sites with a 1–14 bp gap, which we account for as described in *SI Text*, *Prediction of Energy Landscape and Local Diffusion Coefficients*.

Three physical parameters relate *PWM* scores to the effective landscape $U(x)$: PWM_S and c, from Eq. **3**, and the cooperativity term ε, which represents additional binding energy when both dimers are bound in **R** mode. All three parameters are taken from bulk biochemical data (24, 29), described in *SI Text*, *Prediction of Energy Landscape and Local Diffusion Coefficients*.

The effective landscape in Eq. **4** was used to predict local *D* of p53 by calculating over a segment's *n* positions the reduction in diffusivity owing to sequence-specific binding as compared to a featureless landscape, D/D_0, as

$$\frac{D}{D_0} = \left\langle \frac{t_s}{t_{\text{total}}} \right\rangle = \frac{1}{\frac{1}{n}\sum_x^n \exp(-U(x))}. \quad [5]$$

As described in *SI Text*, *Prediction of Energy Landscape and Local Diffusion Coefficients*, we found that D/D_0 in a segment should equal the mean fraction of time a particle in that segment spends sliding, t_s/t_{total}, which in turn equals the reciprocal of the average statistical weight [equivalent to $\exp(-U(x))$] over all sites in the segment of Eq. **4**. We assume that microscopic step rates to positions $x-1$ and $x+1$ are equal.

Once we computed D/D_0 for each segment, we made a correction to account for uncertainty in the assignment of experimental displacements to segments owing to DNA fluctuations, described in *SI Text*, *Interpolations of DNA-Fluctuation Variance and Distributions*. We then assessed the quality of our predicted diffusion coefficients by computing Pearon's correlation coefficient r_{expt} between experimental diffusion coefficients D_{expt} and predicted D/D_0 over the segments, and determined statistical significance of the correlation using a permutation test described in *SI Text*, *Prediction of Energy Landscape and Local Diffusion Coefficients*.

1. Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. *Nature* 408:307–310.
2. Tafvizi A, et al. (2008) Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys J* 95:L01–3.
3. Berg OG, Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–6948.
4. von Hippel PH, Berg OG (1989) Facilitated target location in biological systems. *J Biol Chem* 264:675–678.
5. Kabata H, et al. (1993) Visualization of single molecules of RNA polymerase sliding along DNA. *Science* 262:1561–1563.
6. Harada Y, et al. (1999) Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophys J* 76:709–715.
7. Blainey PC, van Oijen AM, Banerjee A, Verdine GL, Xie XS (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc Natl Acad Sci USA* 103:5752–5757.
8. Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–1194.
9. Wang YR, Austin RH, Cox EC (2006) Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys Rev Lett* 97:049302.
10. Gorman J, et al. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2–Msh6. *Mol Cell* 28:359–370.
11. Gorman J, Greene EC (2008) Visualizing one-dimensional diffusion of proteins along DNA. *Nat Struct Mol Biol* 15:768–774.
12. McKinney K, Mattia M, Gottifredi V, Prives C (2004) p53 linear diffusion along DNA requires its C terminus. *Mol Cell* 16:413–424.
13. Liu Y, Lagowski J, Vanderbeek G, Kulesz-Martin M (2004) Facilitated search for specific genomic targets by p53 C-terminal basic DNA binding domain. *Cancer Biol Ther* 3:1102–1108.
14. Hu T, Grosberg AY, Shklovskii BI (2006) How proteins search for their specific sites on DNA: The role of DNA conformation. *Biophys J* 90:2731–2744.
15. Gowers DM, Wilson GG, Halford SE (2005) Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci USA* 102:15883–15888.
16. Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* 332:475–478.
17. Slutsky M, Mirny LA (2004) Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys J* 87:4021–4035.
18. Mirny L, et al. (2009) How a protein searches for its site on DNA: The mechanism of facilitated diffusion. *J Phys A* 42:434013.
19. Winter RB, Berg OG, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor–operator interaction: Kinetic measurements and conclusions. *Biochemistry* 20:6961–6977.
20. Kalodimos C, et al. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* 305:386–389.
21. Kalodimos C, Boelens R, Kaptein R (2004) Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the lac repressor system. *Chem Rev* 104:3567–3586.
22. Townson SA, Samuelson JC, Bao Y, Xu SY, Aggarwal AK (2007) Bstyi bound to noncognate DNA reveals a "hemispecific" complex: Implications for DNA scanning. *Structure* 15:449–459.
23. Iwahara J, Zweckstetter M, Clore GM (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc Natl Acad Sci USA* 103:15062–15067.
24. Weinberg RL, Veprintsev DB, Fersht AR (2004) Cooperative binding of tetrameric p53 to DNA. *J Mol Biol* 341:1145–1159.
25. Melero R, et al. (2011) Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. *Proc Natl Acad Sci USA* 108:557–562.
26. Tafvizi A, Huang F, Fersht AR, Mirny LA, van Oijen AM (2011) A single-molecule characterization of p53 search on DNA. *Proc Natl Acad Sci USA* 108:563–568.
27. Klein C, et al. (2001) NMR spectroscopy reveals the solution dimerization interface of p53 core domains bound to their consensus DNA. *J Biol Chem* 276:49020–49027.
28. Horvath MM, Wang X, Resnick MA, Bell DA (2007) Divergent evolution of human p53 binding sites: Cell cycle versus apoptosis. *PLoS Genet* 3:e127.
29. Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315:233–237.
30. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys* 22:403–434.
31. Menendez D, Inga A, Resnick MA (2009) The expanding universe of p53 targets. *Nat Rev Cancer* 9:724–737.
32. Zhang C, et al. (2006) A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. *Nucleic Acids Res* 34:2238–2246.
33. Ma B, Pan Y, Zheng J, Levine AJ, Nussinov R (2007) Seque nce analysis of p53 response-elements suggests multiple binding modes of the p53 tetramer to DNA targets. *Nucleic Acids Res* 35:2986–3001.
34. Wei CL, et al. (2006) A global map of p53 transcription-factor binding sites in the human genome. *Cell* 124:207–219.
35. Veprintsev DB, Fersht AR (2008) Algorithm for prediction of tumour suppressor p53 affinity for binding sites in DNA. *Nucleic Acids Res* 36:1589–1598.
36. Jordan JJ, et al. (2008) Noncanonical DNA motifs as transactivation targets by wild type and mutant p53. *PLoS Genet* 4:e1000104.
37. Levy Y, Wolynes P, Onuchic J (2004) Protein topology determines binding mechanism. *Proc Natl Acad Sci USA* 101:511–516.
38. Halford S, Marko J (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* 32:3040–3052.
39. Zhou HX (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc Natl Acad Sci USA* 108:8651–8656.
40. Weinberg RL, Veprintsev DB, Bycroft M, Fersht AR (2005) Comparative binding of p53 to its promoter and DNA recognition elements. *J Mol Biol* 348:589–596.